(1)

ARI Research Note 90-64

The Army Family Research Program: Sampling Plan for the CORE Research Program

Vincent G. Iannacchione
Research Triangle Institute

for

Contracting Officer's Representative
D. Bruce Bell

Personnel Utilization Technical Area
Paul A. Gade, Chief

Manpower and Personnel Research Laboratory
Zita M. Simutis, Acting Director

July 1990

United States Army
Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

JON W. BLADES
COL, IN
Commanding

Research accomplished under contract for the Department of the Army

Research Triangle Institute

Technical review by

Arthur C. F. Gilbert

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS --
2a. SECURITY CLASSIFICATION AUTHORITY -- 2b. DECLASSIFICATION/DOWNGRADING SCHEDULE --	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S) --	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Research Note 90-64
6a. NAME OF PERFORMING ORGANIZATION Research Triangle Institute 6b. OFFICE SYMBOL (If applicable) --	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences
6c. ADDRESS (City, State, and ZIP Code) P.O. Box 12194 Research Triangle Park, NC 27709	7b. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences 8b. OFFICE SYMBOL (If applicable) PERI-R	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA903-87-C-0540
8c. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 63007A PROJECT NO. 792 TASK NO. (2302) 242 WORK UNIT ACCESSION NO. C2

11. TITLE (Include Security Classification)
The Army Family Research Program: Sampling Plan for the CORE Research Program

12. PERSONAL AUTHOR(S)
Iannacchione, Vincent G.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 86/11 TO 88/06	1990, July	34

16. SUPPLEMENTARY NOTATION
Contracting Officer's Representative, D. Bruce Bell

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB-GROUP	Family adaptation	Spouse employment	Leadership
			Retention	Readiness	Policy
			Reenlistment	Community support	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)
The Army Family Research Program (AFRP) is a 5-year integrated research program that supports the Chief of Staff of the Army (CSA) White Paper 1983: The Army Family and The Army Family Action Plans (1984-1990) through the development of databases, models, program evaluation technologies, and policy options that assist the Army to retain quality soldiers, improve soldier and unit readiness, and increase family adaptation to Army life. This report provides details of the sampling design for the soldier portion of a worldwide sample of soldiers, leaders, service providers, and spouses that was conducted between February 1989 and May 1990. Specific topics addressed include (1) a description of the population being sampled, (2) the sampling frames used during each of the three sampling stages, (3) the allocation of types of individuals to be sampled, and (4) the rules for selecting potential participants for the survey.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL D. Bruce Bell	22b. TELEPHONE (Include Area Code) (202) 274-8119 22c. OFFICE SYMBOL PERI-RP

THE ARMY FAMILY RESEARCH PROGRAM: SAMPLING PLAN FOR THE CORE RESEARCH PROGRAM

CONTENTS

LIST OF EXHIBITS

THE ARMY FAMILY RESEARCH PROGRAM: SAMPLING PLAN FOR THE CORE RESEARCH PROGRAM

1. INTRODUCTION

The research objectives of the Army Family Research Program (AFRP) require that characteristics and attitudes of soldiers and their family members be related to characteristics of the Army at both the unit and installation level. For example, outcomes for individuals such as family wellness and soldier readiness must be related to characteristics and outcomes of units, such as unit readiness, leadership attitudes, and unit mission. As a consequence, the sampling design must include provisions for representing the units of the Army and individuals from those units, as well as their spouses if they are married. To support these research objectives, the probability samples of persons and units must produce unbiased estimates of soldier characteristics, characteristics of soldiers' spouses and families, and characteristics of Army units.

The purpose of this report is to provide details of the sampling design that is planned for the core research effort of the AFRP. The report expands and elaborates on the sampling design that was presented in the program's research plan (Barokas & Croan, 1988). Specifically, a description of the data that have been obtained for the construction of the first- and second-stage sampling frames is presented, along with summary tabulations of relevant data elements. Also presented are details of the sampling frames, stratum allocations, and sample selection procedures.

2. OVERVIEW OF THE SAMPLING DESIGN

To be considered valid in a statistical sense, any inferences drawn from a sample must be supported by the probability structure that gave rise to the observations in hand. The underlying probability structure provides the required link between the sample and the survey population. The specification of the probability structure is conveniently referred to as the sample design.

The sample design for this effort can be summarized as a stratified, three-stage, cluster design. Relevant statistical principles have been used to develop a demonstrably unbiased design. The requirements for an unbiased design are that:

- Every member of the survey population be assigned a nonzero probability of selection into the sample, and,
- The randomization procedure used to select the sample generates, in expectation, the assigned probabilities for each member of the survey population.

Given these requirements, specific design issues then center on assigning the probabilities in such a manner as to obtain acceptable levels of precision for acceptable levels of cost.

The survey population is comprised of persons eligible to be selected into the survey sample. For the Army Family Research Program the population is defined as:

All Army personnel at pay-grade levels E2 through O6,
on active duty at the time of data collection,
except persons assigned to classified units,
who are assigned to a permanent duty station
where Army family services are readily accessible, and
all spouses of the above personnel.

The sample will be selected from the survey population in three stages. The first stage of selection will be comprised of geographically proximate sites (usually installations), each of which can be partitioned into second-stage units consisting of deployable/functional organizational units, each of which contains a minimum number of soldiers. A sample of 40 first-stage sampling units (FSUs) will be selected with probabilities proportional to size (PPS) where the size measure is the composite number of eligible soldiers. A composite size measure is used to attain, in expectation, the desired second- and third-stage sample allocations. Within selected FSUs, a PPS sample of 480 second-stage units (SSUs) will be chosen. Finally, a random sample of approximately 17,900 soldiers and their spouses will be selected from the selected SSUs.

Stratification will be used at each stage of selection to control the distribution of the sample with respect to important organizational and demographic characteristics. These include geographic region at the first stage, unit function at the second stage, and demographic categories defined by paygrade, sex, and marital status at the third stage. At each stage, selection probabilities will be assigned to sampling units to yield a self-weighting sample of persons within categories defined by the intersection of unit function and demographic category. The sample design, summarized in Exhibit 1, is described in detail below.

Exhibit 1. Overview of the AFRP Core Research Effort Sample Design

First Stage	
Sampling Units:	Posts/installations/sites
Stratification:	Geographic region
Allocation to Strata:	Proportional to composite number of persons
Type of Selection:	PPS[1] to composite number of persons
Sample Size:	40
Second Stage	
Sampling Units:	Army organizational units (UICs)
Stratification:	Unit function (MTOE, TDA)
Allocation to Strata:	350 MTOE units, 130 TDA units
Type of Selection:	PPS[1] to composite number of persons
Sample Size:	480
Third Stage	
Sampling Units:	Soldiers and spouses of soldiers
Stratification:	Paygrade Group, Sex, and Marital Status
Allocation to Strata:	Oversample officers, marrieds, and females
Type of Selection:	Simple random sample
Initial Sample Size:	Approximately 17,900 soldiers and 14,500 spouses
Final Sample Size[2]:	Approximately 14,400 soldiers and 11,600 spouses

1 Probability proportional to size.
2 Assumes an 80 percent participation rate.

3. DATA SOURCES

The three-stage nature of the sampling design requires that three aspects of Army operational units be delineated: 1) their distribution by geographic location, 2) their size and function, and 3) the characteristics of the persons assigned to them (i.e. paygrade, sex, and marital status). After an indepth examination of Army data sources, it was concluded that data from the Officer and Enlisted Personnel Master Files, maintained by the Total Army Personnel Agency (TAPA) would provide the requisite information.

In May, 1988 a data file was created from the personnel master files that identified the Unit Identification Code (UIC), Army Location Code (ARLOC), major command, and deployable status of all active duty, non-classified operational units in the Army. In addition, counts of the number of persons assigned to these units by paygrade, sex, and marital status were obtained. The file accounts for 714,887 active-duty Army personnel in paygrades E2 through O6 stationed in 1,150 locations throughout the world.

As of this writing, these data have been used to construct the first-stage sampling frame that is described in Section 4.1. After the first-stage sample is selected, these data also will be used to construct the second-stage sampling frame that is described in Section 4.2. The third-stage sampling frame will consist of individuals assigned to selected UICs at the time of data collection. To account for personnel changes that occur between the selection of the second- and third-stage samples, a data file containing randomly selected roster positions will be sent to TAPA one month prior to data collection. The roster positions will be matched with persons on the current version of the master personnel files. Persons identified during the matching process will comprise the third-stage sample. Details of the construction of the third-stage sampling frame are provided in Section 4.3.

4. SAMPLING FRAMES

4.1 First-Stage Sampling Frame

The first-stage sampling frame is comprised of first-stage sampling units (FSUs) that are defined as (a) single geographic sites (i.e. ARLOCs) where Army personnel are located, or (b) a combination of geographic sites. In particular, each FSU is required to have one site (called a nucleus site) that contains at least 1,000 eligible persons. There are two reasons for this requirement: it ensures a cost-effective size for data collection; and, it increases the likelihood that Army support services are provided within the area encompassed by the FSU.

Optionally, an FSU also may contain one or more non-nucleus sites (called satellite sites) that are within 50 miles of the nucleus site. The association of nearby satellite sites (e.g. recruiting stations) with a nucleus site is desirable because personnel at the satellite sites are likely to use the Army support services provided by the nucleus site.

As of this writing, the assignment of eligible personnel to FSUs is complete. A total of 131 locations each contain 1,000 or more eligible persons and have been declared nucleus sites. Persons stationed at nucleus sites account for 618,449 (87%) of the 714,887 persons in paygrades E2 through O6. Persons stationed at 343 satellite sites account for 56,573 (8%) of the total. The remaining 39,865 (5%) persons are more than 50 miles from a nucleus site and have been excluded from the target population. Thus, the target population is comprised of the 675,022 persons stationed at nucleus or satellite sites.

First-stage strata will be defined by geographic region of the world. Details of the first-stage stratum allocations are provided in Section 5.1.

4.2 Second-Stage Sampling Frame

Second-stage sampling units (SSUs) will be defined in terms of Army organizational units identified by one or more UICs and generally will correspond to companies. This definition of SSUs is required, both substantively and analytically, to investigate the issue of unit readiness, because for the Army, a unit is not simply an aggregation of persons but is a discrete functioning entity.

In order to attain the overall sample size of over 17,900 persons, an average of 38 persons must be selected from each of the 480 selected SSUs. However, because some operational units have fewer than 38 persons, the following minimum size requirement will be applied. Each SSU will be comprised of one UIC that has 50 or more individuals assigned to it. Optionally, an SSU may contain one or more other UICs at the same location that contain fewer than 50 persons. The association of UICs within an SSU will follow organizational lines whenever possible.

Second-stage strata will be defined in terms of *unit function* as per the Modified Table of Organization and Equipment (MTOE), and the Table of Distribution and Allowances (TDA). Such stratification will serve to assure representation, to the extent that they exist in sample FSUs, of certain types of units that are critical to the analytical requirements of the survey.

4.3 Third-Stage Sampling Frame

Several months will elapse between the selections of the second- and third-stage samples. To account for changes in the number and distribution of persons that occur during this time, third-stage sampling units will be defined as positions on unit rosters rather than specific individuals. At the time the second-stage sample is selected, positions on a conceptual roster will be serially numbered and a random sample of line numbers

selected. Then, about a month prior to the start of data collection, the sample line numbers will be applied to the actual roster of individuals. If a decrease in a unit's personnel complement has occured since the unit was selected, some of the line numbers will be empty. An increase in personnel will be accommodated by considering the roster to be circular, thereby allowing more than one individual to correspond to the same line number.

An average of 38 soldiers will be randomly selected for inclusion in the survey from each of the 480 selected SSUs, yielding a total sample of approximately 17,900 soldiers and 14,500 spouses. A total of twenty third-stage strata will be formed on the basis of paygrade group, sex, and marital status. The distribution of the target population by paygrade, sex, and marital status is shown in Exhibit 2.

Exhibit 2. Distribution of the Target Population by Paygrade Group, Marital Status, and Sex

Paygrade Group	Marital Status	Males		Females		Total	
		Persons	%	Persons	%	Persons	%
E2-E4	Married	95,267	14.11	15,890	2.35	111,157	16.47
	Not Married	195,154	28.91	25,170	3.73	220,324	32.64
		290,421	43.02	41,060	6.08	331,481	49.11
E5-E9	Married	193,673	28.69	13,880	2.06	207,553	30.75
	Not Married	39,906	5.91	8,222	1.22	48,128	7.13
		233,579	34.60	22,102	3.28	255,681	37.88
W1-W4	Married	11,759	1.74	184	0.03	11,943	1.77
	Not Married	1,574	0.23	144	0.02	1,718	0.25
		13,333	1.97	328	0.05	13,661	2.02
01-03	Married	27,130	4.02	3,621	0.54	30,751	4.56
	Not Married	13,981	2.07	3,998	0.59	17,979	2.66
		41,111	6.09	7,619	1.13	48,730	7.22
04-06	Married	21,791	3.23	1,091	0.16	22,882	3.39
	Not Married	1,803	0.27	784	0.12	2,587	0.38
		23,594	3.50	1,875	0.28	25,469	3.77
All	Married	349,620	51.79	34,666	5.14	384,286	56.93
	Not Married	252,418	37.39	38,318	5.68	290,736	43.07
		602,038	89.18	72,984	10.82	675,022	100.00

Source: May, 1988 Officer and Enlisted Personnel Master Files

5. SAMPLE ALLOCATION AND SELECTION

5.1 Sample Allocation

Stratification will be used at each stage of selection to control the distribution of the sample with respect to important geographic, organizational and demographic characteristics. In addition, the probabilities used to select the first- and second-stage samples will be made proportional to a composite size measure to insure that the desired second- and third-stage sample allocations are achieved, in expectation. Details of the proposed composite size measure are presented in the following section.

Geographic region of the world will be used to define three first-stage strata. The number of FSUs to be selected from each stratum is shown in Exhibit 3 and is proportional to the sum of the FSU-specific composite size measures. In turn, the sum of the composite size measures correspond to the expected third-stage sample size for each strata.

Because most of the data analyses planned for the core study will take place at the unit and individual level, the second- and third-stage sample allocations are designed to meet the minimum precision requirements established for both levels of analysis. The third-stage sample allocations are based on the univariate and multivariate statistical analyses planned for the survey. The standard error of an estimated proportion is used to evaluate univariate sample size requirements, while the statistical power associated with estimating the parameters of a multiple regression model is used to determine multivariate sample size requirements for each of the following major subpopulations: males, females, marrieds, not marrieds, E2-E4s, E5-E9s, 01-03s, and 04-06s. Details of the precision requirements are presented in the Appendix of the AFRP Research Plan (Barokas and Croan, 1988).

Exhibit 3. First-Stage Sample Allocation

First-Stage Stratum	Eligible Persons	%	Composite Size[1]	%	Sample Allocation[2]
CONUS, Alaska, Hawaii	437,164	64.8	11,453	63.8	25
Europe	192,602	28.5	5,258	29.3	12
Japan, Korea, Panama, Puerto Rico	45,256	6.7	1,234	6.9	3
	675,022	100.0	17,945	100.0	40

[1]The composite size corresponds to the expected third-stage sample size.

[2]The sample allocation is proportional to the composite size.

The expected precision for univariate and multivariate analyses are presented, for different effective sample sizes, in Exhibit 4. As might be expected, the sample sizes for univariate statistics with acceptable precision are smaller than those for a multiple regression model. For example, an effective sample size of 500 is needed to estimate a percentage to within five percent of the population value with 95 percent confidence. However, an effective sample size of 1,000 is needed to detect relatively large regression parameters with 93 percent power. The results of the power analysis indicate that an effective sample size (i.e. a sample size adjusted for nonresponse and unequal weighting) of at least 1,400 persons is needed for each major subpopulation to detect moderate-sized regression parameters with a statistical power of 0.80.

Unit readiness will be one of the major outcome variables analysed in the research effort. And, because of their mission, deployable (i.e. MTOE) units will be the focus of this analysis. Thus, the minimum precision requirement at the unit level is described in terms of the number of MTOE units allocated. In his April 20 memorandum to the AFRP Core Group (presented in Appendix A), Bob Sadacca, the readiness research area leader, stated that a second-stage sample of at least 350 MTOE units would be needed to detect multiple regression parameters that increase the R^2 value by 0.01 or more with a statistical power of about 0.60.

Because of demographic differences in the composition of MTOE and TDA units, the maximum effective sample size for the proposed third-stage sample allocation occurs with a second-stage sample allocation of 300 MTOE and 180 TDA units. Therefore, an evaluation was made to determine if a disproportionate allocation of 350 MTOE units would cause the effective sample sizes of one or more of the major subpopulations to be less than 1,400. The results of the evaluation, presented in Appendix B, indicate that a second-stage allocation of 350 MTOE units and 130 TDA units, and a third-stage allocation of 17,945 soldiers will satisfy the minimum precision requirements for both the unit-level and person-level analyses. Larger MTOE allocations result in unacceptable precision levels for females and field-grade officers, and cause substantial reductions in the effective

Exhibit 4. Expected Precision for Univariate and Multivariate Statistical Analyses

Effective Sample Size[1]	Regression Parameter[2]			Percent[3]		
	.05	.10	.15	10 or 90	25 or 75	50
				*	*	*
250	.13	.28	.47	3.8	5.4	6.4
500	.19	.44	.72	2.6	3.8	4.4
750	.23	.57	.86	2.2	3.2	3.6
1,000	.28	.68	.93	1.8	2.8	3.2
1,250	.31	.76	.97	1.6	2.4	2.8
1,500	.36	.82	.99	1.5	2.2	2.6
1,750	.40	.87	.99+	1.4	2.0	2.4
2,000	.44	.91	.99+	1.3	1.9	2.2
2,250	.47	.93	.99+	1.3	1.8	2.1
2,500	.49	.95	.99+	1.2	1.7	2.0
2,750	.54	.97	.99+	1.1	1.7	1.9
3,000	.57	.98	.99+	1.1	1.6	1.8

[1]The effective sample size is obtained by multiplying the initial sample size by the response rate and then dividing by the unequal weighting effect.

[2]Values are estimates of statistical power for a one-tailed test with α=.05. For example, regression models with an effective sample size of 1,000 will detect parameters of size .10 or larger 68 percent of the time.

[3]Values are 95 percent confidence intervals. For example, percentages based on an an effective sample size of 750 will be within 3.6 percent of the population value 95 percent of the time.

sizes of the other subpopulations. The number of SSUs to be selected from each first- and second-stage stratum combination is shown in Exhibit 5 and is proportional to the SSU-specific composite size measure described in the following section.

Military personnel will be classified into twenty third-stage strata determined by the intersection of sex, marital status (i.e. married, not married), and paygrade group (i.e. E2-E5, E6-E9, W1-W4, 01-03, and 04-06). In order to meet the analytic requirements of the survey, the third-stage sample will be comprised of greater proportions of officers, marrieds, and females than exist in the target population. The initial and effective distribution of sample persons across these strata, and the unequal weighting effects caused by disproportionate sampling are shown in Exhibit 6.

5.2 Composite Size Measures

Composite size measures will be used at the first and second-stages of sample selection to insure that the targeted sample sizes are achieved, in expectation, for the subpopulations of interest, (i.e. the third-stage strata). The composite size measures will be formulated in the following manner. Let $c = 1,2,\ldots,20$ index the subpopulations of interest shown in Exhibit 6, and let n_c designate the desired sample size for subpopulation c. Further, define $N_c(i,j)$ as the number of eligible soldiers in SSU j of FSU i that belong to subpopulation c. Now define the following person-level frame counts:

$$N_c(i) = \sum_j N_c(i,j), \text{ and,}$$

$$N_c = \sum_i \sum_j N_c(i,j).$$

Thus, the desired sampling rate for members of subpopulation c is

$$f_c = n_c / N_c.$$

Exhibit 5. Second-Stage Sample Allocation

First-Stage Stratum	Second-Stage Stratum	Eligible Persons	%	Composite Size[1]	%	Sample Allocation[2]
CONUS, Alaska, Hawaii	MTOE	261,623	38.8	7,176	40.0	192
	TDA	175,541	26.0	4,277	23.8	114
		437,164	64.8	11,453	63.8	306
Europe	MTOE	176,588	26.2	4,872	27.1	130
	TDA	16,014	2.4	386	2.2	11
		192,602	28.5	5,258	29.3	141
Japan, Korea, Panama, Puerto Rico	MTOE	37,578	5.6	1,036	5.8	28
	TDA	7,678	1.1	198	1.1	5
		45,256	6.7	1,234	6.9	33
Overall	MTOE	475,789	70.5	13,084	72.9	350
	TDA	199,233	29.5	4,861	27.1	130
		675,022	100.0	17,945	100.0	480

[1]The composite size corresponds to the expected third-stage sample size.

[2]The sample allocation is proportional to the composite size.

Exhibit 6. Third-Stage Sample Allocation Given a Second-Stage Allocation of 350 MTOE and 130 TDA SSUs.

Paygrade Group	Marital Status	Male			Female			Total		
		Initial Sample Size	Effective Sample Size*	Unequal Weighting Effect**	Initial Sample Size	Effective Sample Size*	Unequal Weighting Effect**	Initial Sample Size	Effective Sample Size*	Unequal Weighting Effect**
E2-E4	Married	5,440	4,271	1.02	1,277	1,015	1.01	6,717	5,205	1.03
	Not Mar	1,360	1,068	1.02	340	264	1.03	1,700	1,276	1.07
		-----			-----			-----		
	Subtot	6,800	2,232	2.44	1,617	638	2.03	8,417	2,718	2.48
E5-E9	Married	2,624	2,045	1.03	656	510	1.03	3,280	2,302	1.14
	Not Mar	656	512	1.02	164	127	1.03	820	636	1.03
		-----			-----			-----		
	Subtot	3,280	2,544	1.03	820	538	1.22	4,100	2,924	1.12
W1-W4	Married	256	200	1.02	15	12	1.01	271	205	1.06
	Not Mar	64	50	1.02	12	9	1.01	76	57	1.06
		-----			-----			-----		
	Subtot	320	240	1.07	27	21	1.01	347	250	1.11
O1-O3	Married	2,169	1,723	1.01	284	225	1.01	2,453	1,949	1.01
	Not Mar	560	437	1.03	140	109	1.03	700	544	1.03
		-----			-----			-----		
	Subtot	2,729	1,932	1.13	424	283	1.20	3,153	2,199	1.15
O4-O6	Married	1,653	1,315	1.01	82	65	1.00	1,735	1,380	1.01
	Not Mar	136	108	1.01	59	47	1.00	195	155	1.00
		-----			-----			-----		
	Subtot	1,789	1,423	1.01	140	112	1.00	1,929	1,535	1.01
ALL	Married	12,142	5,570	1.74	2,313	1,702	1.09	14,455	6,520	1.77
	Not Mar	2,776	1,621	1.37	714	475	1.20	3,490	1,994	1.40
		------			------			-----		
	Total	14,918	5,918	2.02	3,027	1,404	1.73	17,945	7,006	2.05

Source: May,1988 Officer and Enlisted Personnel Master Files.

* The effective sample size is the initial sample size multiplied by the response rate (assumed to be 80%) and then divided by the unequal weighting effect. It is not additive across strata.

** The unequal weighting effect is the ratio of the variance of a random sample selected with unequal probabilities to the variance of a simple random sample of the same size.

If all SSUs in the population were to be sampled, the sample size of individuals to be selected from subpopulation c in each SSU j of FSU i would be

$$n_c(i,j) = f_c \bullet N_c(i,j).$$

This quantity is the basis for the second-stage composite size measure

$$S(i,j) = \sum_c f_c \bullet N_c(i,j),$$

which may be considered the sample size that would be obtained from SSU j of FSU i if all the SSUs in the population were sampled with the specified sampling rates of f_c for the individual subpopulations. The population total of this size measure is

$$S = \sum_i \sum_j S(i,j)$$

$$= \sum_c f_c \bullet N_c$$

$$= \sum_c n_c ,$$

which is the total third-stage sample size.

The population totals by type of unit are

$$S(MTOE) = \sum_i \sum_{j \epsilon MTOE} S(i,j), \text{ and}$$

$$S(TDA) = \sum_i \sum_{j \epsilon TDA} S(i,j) .$$

If the second-stage sample of 480 units is proportionally allocated on the basis of the above sums, approximately 300 MTOE units will be selected. The allocation of MTOE units can be changed by applying the multiplicative factors, f(MTOE), to the composite size measure of each MTOE unit, and f(TDA), to the composite size measure of each TDA unit.

The second-stage sample allocations to the MTOE and TDA strata then can be expressed as

$$n(\text{MTOE}) = \frac{f(\text{MTOE}) \bullet S(\text{MTOE})}{S} \bullet 480,$$

rounded to the nearest integer, and,

$$n(\text{TDA}) = 480 - n(\text{MTOE}).$$

Thus, for a desired allocation of n(MTOE) units, the multiplicative factors are

$$f(\text{MTOE}) = \frac{n(\text{MTOE}) \bullet S}{S(\text{MTOE}) \bullet 480}, \text{ and}$$

$$f(\text{TDA}) = \frac{n(\text{TDA}) \bullet S}{S(\text{TDA}) \bullet 480}.$$

The adjusted composite size measure assigned to SSU j of FSU i is

$$S'(i,j) = f(\text{MTOE}) \bullet S(i,j), \text{ if } j \epsilon \text{MTOE},$$

and $$S'(i,j) = f(\text{TDA}) \bullet S(i,j), \text{ if } j \epsilon \text{TDA}.$$

Similarly, the adjusted composite measure assigned to an FSU i is

$$S'(i) = \sum_j S'(i,j).$$

First-stage stratum allocations will be made proportional to the sum of the adjusted composite size measures.

An average of 38 eligible soldiers will be selected from each of the 480 selected SSUs. Each of these samples will be allocated to the subpopulations using the adjusted composite size measure assigned to the SSU. The desired allocation to subpopulation c for SSU j in FSU i is

$$n_c(i,j) = 38 \bullet f_c \bullet f(\text{MTOE}) \bullet N_c(i,j) \,/\, S'(i,j), \text{ if } j \epsilon \text{MTOE},$$

and $$n_c(i,j) = 38 \bullet f_c \bullet f(\text{TDA}) \bullet N_c(i,j) \,/\, S'(i,j), \text{ if } j \epsilon \text{TDA}.$$

The desired sampling rate, or third-stage selection probability, for each member of subpopulation c in SSU j of FSU i is

$$f_c(i,j) = 38 \bullet f_c \;/\; S(i,j).$$

The expected sample size for a subpopulation c for a randomly selected SSU j within any randomly selected FSU i can be shown to be

$$E[n_c(i,j)] = f_c \bullet N_c(i,j).$$

Similarly, the expected total sample size for a subpopulation c is

$$\begin{aligned} E[n_c] &= E[\sum_i \sum_j n_c(i,j)] \\ &= \sum_i \sum_j f_c \bullet N_c(i,j) \\ &= f_c \bullet N_c \end{aligned}$$

Thus, the desired sampled sample size, n_c, is achieved for subpopulation c in expectation, or on the average over all possible samples generated by the sample design.

5.3 Sample Selection Procedures

A probability sample of 40 FSUs will be selected with probabilities proportional to the composite number of eligible soldiers in the FSU. Because the five largest sites each account for more than 2.5 percent of the total composite size measure, their selection frequency exceeds one and their entry into the sample is assured. The two largest sites, Ft. Bragg and Ft. Hood, have selection frequencies that exceed two. Thus, each will account for either two or three of the 40 selections.

Except for FSUs that are selected more than once, twelve SSUs will be selected from a selected FSU whenever possible. FSUs that are selected two or three times will have 24 or 36 SSUs selected from them respectively. In the event that one or more FSUs with fewer than twelve SSUs are selected, the allocation of SSUs to other FSUs will be increased to insure an overall sample size of 480 SSUs.

Because FSUs and SSUs will vary considerably with respect to numbers of personnel, the sample will be chosen with minimum replacement (Chromy 1979). The minimum replacement procedure is equivalent to without replacement selection if none of the expected selection frequencies exceed unity, i.e. if there are no self-representing FSUs. Otherwise, the procedure achieves the required frequencies over repeated samples and, at any specific drawing of the sample, comes within one unit of the expected allocation. The minimum replacement method is superior to either with or without replacement schemes in that it controls the number of selections assigned to a sampling unit so that the actual allocation and the proportional-to-size allocation differ by less than one and, at the same time, includes self-representing FSUs with their required frequencies.

To achieve the targeted sample sizes it will be necessary for all selected sites and selected units at those sites to participate in the study. In the event that higher priority activities preclude participation by a site or unit (e.g. IG inspection or unit TDY), a replacement site or unit will be selected in a manner consistent with the sample design.

The replacement strategy for FSUs is to increase the U.S. first-stage stratum allocation by six and the other first-stage stratum allocations by two each with the provision that the additional FSUs be randomly designated as replacements for each stratum. In the event that an FSU is not available to participate in the study, its stratum-specific replacement will be used.

A similar replacement strategy will be used at the second stage. Within an FSU, each second-stage stratum allocation will be increased by

twenty percent or two SSUs, whichever is larger. Then, designated replacements will be randomly assigned from the sample of SSUs. For example, if the second-stage sample allocated to an FSU consists of nine MTOE SSUs and three TDA SSUs, a total of eleven SSUs will be selected from the MTOE stratum and five SSUs will be selected from the TDA stratum. In the event that one of the MTOE SSUs is not available, its designated replacement in the MTOE stratum will be used instead.

At the third stage of selection, every effort will be made to obtain the cooperation of selected individuals. Specifically, a two-phased data collection strategy is proposed to take advantage of the extended site stays planned for the data collection teams. The first phase will occur shortly after a team arrives at a site. Then, to obtain responses from those who unable to attend the first session because of illness, leave, or TDY, a second session will be held before the team leaves the site about two weeks later. This strategy should result in the attainment of the 80 percent response rate that was assumed in determining the targeted sample sizes.

6. REFERENCES

Barokas, J., & Croan, G. (1988). The Army Family Research Program's family factors in retention, readiness, and sense of community: The plan for research. Research Triangle Park, NC: Research Triangle Institute.

Chromy, J. (1979). Sequential sampling selection methods. In Proceedings of the American Statistical Association. Washington, DC: American Statistical Association.

APPENDIX A

POWER ANALYSIS OF MTOE UNIT SMAPLE SIZE

MEMO TO: AFRP Core Group
FROM: Bob Sadacca
SUBJECT: Power Analysis of MTOE Unit Sample Size
DATE: 20 April 1988

In arriving at the requisite sample size for MTOE units below, the following assumptions were made in the power analysis:

1. The sample design effect for the MTOE units will be less than for the individual soldiers (the cluster sampling will be partially offset by stratification of the MTOE units (e.g., by unit type). Let design effect = 1.25.

2. There will be a 5% participation failure rate for individual units once scheduled due to last minute unexpected involvement of the units elsewhere.

3. The number of relatively noncolinear independent variables obtained through separate analyses not involving the dependent variable(s) will 25.

4. R^2 between the independent and dependent variables will be approximately .50.

5. We wish to be able to detect as statistically significant at the .05 level any variable's additional contribution in the regression equation(s) to R^2 that is equal to or greater than .01.

Power Analysis of Semipartial Correlations
(Assumes R^2=.50; r^2_s=.01; independent variables=25)

Power	Approximate Unadjusted Sample Size	Approximate Sample Size Adjusted for Design Effects (1.25) and Participation Failure Rate (.05)
.10	48	62
.30	129	170
.50	218	287
.60	271	357
.70	324	440
.80	418	551
.90	552	726

It can be seen in the above table that unit sample sizes less than 350 will produce probabilities close to the 50/50 point of rejecting false null hypotheses concerning the additional proportion of the dependent variable variance accounted for by any particular independent variable. Obviously, greater power (e.g., .70) would be more desirable as would be the ability to identify r^2_s values less than .01 (an r^2_s of .01 corresponds to a z-score regression or Beta weight of approximately .14, assuming $R^2=.50$). Therefore, a sample of 350 MTOE units can be looked upon as a minimum. If, however, the design and participation effects are somewhat greater than 1.32 or 1.25/.95 (which is admittedly a guess) then a sample size closer to 400 or more MTOE units would be desirable.

APPENDIX B

SECOND AND THIRD STAGE SAMPLE ALLOCATIONS

R E S E A R C H T R I A N G L E I N S T I T U T E

Center for Research in Statistics

MEMORANDUM May 19, 1988

TO: Al Cruze
Bob Sadacca

FROM: Vince Iannacchione VGI

SUBJECT: Second and Third Stage Sample Allocations for the Core Study

The acquisition of the latest data tape from TAPA has made it possible to evaluate the effects of the second-stage sample allocation on the initial and effective sample sizes of the third-stage subpopulations. For expediency, the population used to make this evaluation is slightly larger than the actual target population in that persons located outside the 50 mile radius of nucleus sites are included.

The allocation problem may be summarized as follows. Determine the second- and third-stage sample allocations that will satisfy the minimum precision requirements established for the person-level and unit-level analyses. At our April 27 senior management meeting, I described the person-level precision requirements in terms of effective sample size. In particular, I stated that an effective sample size of at least 1,400 persons would be needed to detect moderate-sized multiple regression parameters with a statistical power of 0.80. In turn, Bob Sadacca stated that a second-stage sample of at least 350 MTOE units would be needed to detect multiple regression parameters that increase the R^2 value by 0.01 or more with a statistical power of about 0.60.

The maximum effective sample size for a given third-stage sample allocation occurs when the second-stage sample is allocated proportional to the sum of the composite size measures described in the sampling plan. For a 480 unit total, a proportional allocation is about 300 MTOE units. For a 400 unit total, a proportional allocation is about 250 MTOE units. Thus, the allocation problem is to determine if a disproportionate allocation of 350 MTOE units causes the effective sample sizes of one or more major subpopulations to be less than 1,400.

The effects of oversampling fall into two categories: unequal weighting caused by selecting MTOE units with higher probability than TDA units of the similar size and composition, and the limitations placed on the composition of the third-stage sampling frame caused by a disproportionate number of MTOE units. For the range of MTOE allocations under consideration, the latter outweighs the former. As can be seen in Exhibit A, the differences in the composition of MTOE and TDA units is quite pronounced for field-grade officers and females. Because of this, the oversampling of these subpopulations is constrained by the number of

MTOE units allocated. For example, if 400 MTOE units and 80 TDA units are allocated, the expected maximum sample size of field-grade officers (04-06) that can be selected is 400x1.7 or 680 from MTOE units, and 80x12.6 or 1,008 from TDA units. Thus, 1,688 field-grades is the largest initial sample size that can be obtained (if no other subpopulations are considered) with this allocation.

The initial and effective sample sizes of the major subpopulations for second-stage sample allocations of 480 units and 400 units are presented in Exhibits B and C respectively. In each case, changes to the MTOE allocation were affected by applying a multiplicative factor (presented in footnote 2 of each Exhibit) to the composite size measures of MTOE units. For example, to obtain 350 MTOE units from a 480 unit total, the composite size measure of every MTOE unit on the sampling frame was multiplied by 1.61. As a result, the composite size measure of an MTOE unit would be 1.61 times as large as a TDA unit with the same number of persons and subpopulation distribution.

The implications of this evaluation are clear:

- For a 480 unit total, an allocation of 350 MTOE units will satisfy the minimum power criteria for both the unit-level and person-level analyses. Larger MTOE allocations result in unacceptable precision levels for females and field-grade officers and cause substantial reductions in the effective sizes of the other subpopulations.

- For a 400 unit total, it is not possible to satisfy both minimum power criteria simultaneously. An allocation of 350 MTOE units results in unacceptable levels of precision for females and field-grade officers. Further, to obtain a sample that is 87% MTOE, a multiplicative factor of 3.71 must be applied to the composite size measure of every MTOE unit. This produces an unequal weighting effect that substantially reduces the effective sample size of every subpopulation. Alternatively, an allocation of 300 MTOE units out of 400 total units will satisfy the minimum power criteria for the person-level analyses, but will reduce the probability of detecting 0.01 changes in the R^2 of unit-level multiple regressions from 0.60 to 0.51.

Attachments

cc: Ralph Folsom
    Jenny Milne

Exhibit A. Mean Per Unit Distribution of Subpopulations by Type of Unit[1]

Subpopulation	MTOE Mean	MTOE %	TDA Mean	TDA %
<u>Paygrade Group</u>				
E2 - E4	64.3	54.2	43.3	34.1
E5 - E9	42.8	36.1	55.7	43.8
W1 - W4	2.4	2.0	2.5	2.0
01 - 03	7.3	6.2	12.9	10.1
04 - 06	1.7	1.4	12.6	9.9
<u>Sex</u>				
Male	108.1	91.1	107.6	84.7
Female	10.5	8.9	19.5	15.3
<u>Marital Status</u>				
Married	64.1	54.0	82.9	65.2
Not Married	54.4	46.0	44.2	34.8
<u>Overall</u>				
	118.6	100.0	127.1	100.0

[1]Assumes a total of 4,100 MTOE units and 1,800 TDA units.

Exhibit B. Initial and Effective Sample Sizes of Subpopulations for a Second-Stage Sample of 480 Units[1]

Sub-Population	Desired Initial Sample Size	Number of MTOE Units Allocated[2]					
		300		350		400	
		Initial	Effect.	Initial	Effect.	Initial	Effect.
Paygrade Group[3]							
E2-E4	8,500	8,500	2,784	8,474	2,683	8,415	2,323
E5-E9	4,100	4,100	3,002	4,100	2,853	4,100	2,364
01-03	3,500	3,292	2,305	3,251	2,210	3,162	1,850
04-06	3,500	2,641	2,088	2,244	1,786	1,701	1,217
Sex							
Male	16,000	15,640	6,182	15,286	5,925	14,740	5,052
Female	4,000	3,243	1,470	3,131	1,394	2,987	1,142
Marital Status							
Married	16,000	15,302	6,691	14,895	6,361	14,263	5,273
Not Married	4,000	3,581	2,066	3,522	1,988	3,463	1,717
Overall							
	20,000	18,883	7,327	18,417	7,019	17,727	5,970

Source: May, 1988 Officer and Enlisted Personnel Master Files (TAPA).

[1]The effective sample size is the initial sample size multiplied by the response rate (assumed to be 80%) and the divided by the unequal weighting effect.

[2]The following multiplicative factors were applied to the composite size measures of MTOE units: 1.03 to obtain 300 MTOE units,
1.61 to obtain 350 MTOE units, and
2.82 to obtain 400 MTOE units.

[3]Warrant officers are not shown, but are sampled at rate proportional to their occurance in the population.

Exhibit C. Initial and Effective Sample Sizes of Subpopulations for a Second-Stage Sample of 400 Units[1]

Sub-Population	Desired Initial Sample Size	Number of MTOE Units Allocated[2]					
		250		300		350	
		Initial	Effect.	Initial	Effect.	Initial	Effect.
Paygrade Group[3]							
E2-E4	8,500	8,301	2,782	8,242	2,656	8,183	2,093
E5-E9	4,100	4,100	3,003	4,100	2,819	4,100	2,066
01-03	3,500	3,112	2,211	2,896	2,050	2,680	1,450
04-06	3,500	2,322	1,825	1,779	1,404	1,237	789
Sex							
Male	16,000	15,232	6,174	14,559	5,850	13,885	4,477
Female	4,000	2,947	1,448	2,802	1,352	2,657	971
Marital Status							
Married	16,000	14,647	6,652	13,888	6,227	13,129	4,550
Not Married	4,000	3,531	2,066	3,472	1,969	3,413	1,540
Overall							
	20,000	18,179	7,311	17,361	6,922	16,542	5,276

Source: May, 1988 Officer and Enlisted Personnel Master Files (TAPA).

[1]The effective sample size is the initial sample size multiplied by the response rate (assumed to be 80%) and the divided by the unequal weighting effect.

[2]The following multiplicative factors were applied to the composite size measures of MTOE units: 0.98 to obtain 250 MTOE units,
1.69 to obtain 300 MTOE units, and
3.71 to obtain 350 MTOE units.

[3]Warrant officers are not shown, but are sampled at rate proportional to their occurance in the population.